Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *The Effectiveness of Disease-modifying Anti-rheumatic drugs (DMARDs) in Children with Juvenile Idiopathic Arthritis (JIA)*

Draft review available for public comment from November 15, 2010 to December 13, 2010.

Research Review Citation: Kemper AR, Coeytaux R, Sanders GD, Van Mater H, Williams JW, Gray RN, Irvine RJ, Kendrick A. The Effectiveness of Disease-modifying Anti-rheumatic drugs (DMARDs) in Children with Juvenile Idiopathic Arthritis (JIA.) Comparative Effectiveness Review No. 28. (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-02-0025.) AHRQ Publication No. 11-EHC039-EF. Rockville, MD: Agency for Healthcare Research and Quality. September 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or E-mail. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Executive Summary	JIA has “categories” of disease not subtypes. Please make this change throughout the document. Further there are 7 categories: Systemic, oligo, RF-poly RF+poly, psoriatic, enthesitis related, and undifferentiated. Please see attached paper (Petty, Southwood, Manners, et al., Journal of Rheumatology 2004;31:2, 390-2)	This change has been made.
Peer Reviewer #2	Executive Summary (Pg 8, line 40)	Please put the word “spiking” before “fever”	This change has been made.
Peer Reviewer #2	Executive Summary (Pg 9, line 26)	I believe that the FDA warnings are referred to as “box” warnings, rather than “black box” warnings	This change has been made.
Peer Reviewer #2	Executive Summary (Pg 10, line 32)	I would suggest clarifying which meds are intra-articular	Only the corticosteroids are intra-articular. The text specifically refers to “intra-articular corticosteroids.”
Peer Reviewer #2	Executive Summary (KQ 4)	Again – categories of JIA; RF- and RF+ are separate	This change has been made.
Peer Reviewer #2	Executive Summary	Table A is excellent	Thank you.
Peer Reviewer #2	Executive Summary	Information regarding the amount and quality of the information available regarding treatment of JIA compared to other chronic diseases of childhood that occur with similar frequency would be helpful.	This review focuses only on the treatment of JIA. Therefore, we do not compare the amount and quality of evidence to the treatment of other rare but important childhood illnesses.
Peer Reviewer #2	Executive Summary (Pg 12, line 14)	I would suggest a notation (or footnote) that explains WHY the ESR is inconsistently associated with treatment. It is because many (?Majority) of children with active JIA do not have an elevated ESR – so there is no room for it to improve with treatment.	This concern has been addressed under “Remaining Issues.”
Peer Reviewer #2	Executive Summary	Likewise with regard to lack of radiographic data – it might be helpful to have a notation regarding lack of standardization and difficulty reading films on joints that still have so much cartilage.	This concern has been addressed under “Remaining Issues.”
Peer Reviewer #2	Executive Summary	With regard to health status – it would be helpful to have a notation that the instrument most often used – CHAQ – does not do a good job of capturing how kids are really functioning.	Explanation of CHAQ as disability index has been clarified.
Peer Reviewer #2	Executive Summary	This might also be a good place to mention that investigations should also report disease states, not just “improvement”.	This has been added.
Peer Reviewer #2	Executive Summary (Pg 15, line 42)	“Macrophage” rather than macrocyte.	This change was made.

Peer Reviewer #2	Executive Summary (Pg 15, line 46 & 56)	A disease registry rather than a DMARD registry is key for evaluation of risk of JIS treatments.	This is now described in the text.
Peer Reviewer #6	Executive Summary	Executive summary - page 8 last bulleted item on page. "Macrocyte" should be "macrophage".	This change was made.
Peer Reviewer #6	Executive Summary	In general, excellent. On page 16, consider inclusion of text to say that the FDA specifically identified hepatosplenic lymphoma and that interpretation of results is difficult because of inclusion of IBD patients, concurrent medications, and assumption that incidence of lymphoma among JIA patients is no different than that of normal children.	We appreciate this comment. Our understanding is that the FDA based this early announcement in 2008 on data obtained in part from adult patients, and that there is currently a fair amount of controversy surrounding this topic. This issue was not identified by the eligible literature identified at the time that we prepared the draft report. Upon updating our literature search in preparation for the final report, however, we identified two recently published studies that investigate possible relationships between TNFα blockers and cancer, particularly lymphoma. We discuss these studies and their findings in the KQ3 sections of the revised report.
Miller, Amy S.	Executive Summary	The American College of Rheumatologyappreciates AHRQ's attention to JIA in the form of this project. Clearly, more research is needed in this field, and the report is very helpful in identifying the numerous research gaps. The ACR engourages AHRQ to go one step further, and fund and disseminate Requests for Proposals for this type of research, particularly in the area of JIA clinical outcome measures.	This report has made AHRQ aware of the important gaps in knowledge, and we hope that this will serve as the basis for new research.
Olson, Judyann	Executive Summary	On page 8, there seems to be an oversight – there is mention of uveitis with many of the subtypes of JIA, but not poly – while the incidence is lower in this subtype, uveitis still occurs.	This is now described.
Peer Reviewer #1	Introduction	ILAR classified JIA into 7 "categories", not "subtypes". This is a common mistake throughout our literature, but to be accurate the text should be changed to "categories" whenever referencing the ILAR system. If referring to JIA phenotypes without specific reference to ILAR, then fine to use "subtypes", "phenotypes", or any other tem preferred by the authors.	We now use categories throughout.

Peer Reviewer #1	Introduction	Would make more clear the distinction between biologics and non-biologics – the difference is not simply route of administration as implied by the text. In fact, some biologics in development are oral medications. Biologics are created by biologic processes, as opposed to chemical manufacturing such as for methotrexate. The newer biologics (essentially excluding IVIG) are specifically targeted at identified components of the immune system, such as signaling or cell-surface molecules, as opposed to methotrexate for which the mechanisms of action are not known. (I'm sure you can find a better definition of biologics somewhere, but this is the gist…)	This has now been strengthened, following the recommendation.
Peer Reviewer #2	Introduction	JIA has "categories" of disease not subtypes. Please make this change throughout the document. Further there are 7 categories: Systemic, oligo, RF-poly RF+poly, psoriatic, enthesitis related, and undifferentiated. Please see attached paper (Petty, Southwood, Manners, et al., Journal of Rheumatology 2004;31:2, 390-2)	This has been corrected.
Peer Reviewer #2	Introduction (Pg 17, line 44)	Long term complications also include death, blindness, osteoporosis, organ involvement, social problems, persistently active disease and unemployment.	More examples of long-term complications are now included.
Peer Reviewer #2	Introduction (Pg 19, line 54)	RF-poly and RF+poly - these really are two different diseases.	This change has been made.
Peer Reviewer #2	Introduction	Table 1 – I may be quite wrong, but I thought that Arava was FDA approved for JIA? I did not think that Indocin ever received FDA approval.	Leflunomide is "indicated in adults for the treatment of active rheumatoid arthritis (RA)." The label describes the pediatric studies but does not state approval. The FDA has not specifically approved indomethacin for JIA. This has been corrected.
Peer Reviewer #3	Introduction	The introduction provides a comprehensive overview of the topic and the rationale for conducting this CER. Table 1 could be improved by adding a column with recommended dosing intervals and modes of application for the different drugs. Particularly for biologics, they vary substantially among drugs and often play an important role in choosing a medication	This review is not intended for use by clinicians to provide treatment. Providing this level of dosing information goes beyond our evaluation.

Peer Reviewer #4	Introduction	The target population is explicitly defined, and the limitations of defining the target population in JIA are nicely discussed elsewhere in the report. As a clinician, I want to know how best to treat an individual patient with a defined subtype of JIA (KQ4). While I do not think that key questions 1 and 2 are clinically meaningful, they ARE APPROPRIATE given that the answer to the most meaningful key question, #4, is not obtainable due to lack of data. Yes, the key questions are appropriate and explicitly stated. Should there be a key question about treating JIA-associated uveitis? Probably not, but that is a problem that is unlikely to be considered significant enough to have its own comparative effectiveness review. So why not here?	The key questions were informed and developed through a separate topic development process. We are unable to modify the key questions.
Peer Reviewer #4	Introduction	Why are there no studies of gold included (p.19 list of DMARDs) (aside from the one mentioned on p46.)? Gold is no longer in routine use for JIA, but neither is Penicillamine, or several of the other DMARDs listed. For the sake of thoroughness, gold should be included in the DMARD list and studies including gold should be included in the analysis if they meet inclusion criteria.	We included studies of gold if they met the described inclusion criteria. No other studies met the criteria.
Peer Reviewer #4	Introduction	I believe there is a typo in Table 2, on pg 22 – Kenolog should be Kenalog. I noticed this on another table in the report, but I do not recall where.	Correction has been made.
Peer Reviewer #5	Introduction	The introduction explains the background very thoroughly and appropriately. Key questions are well defined and clinically meaningful. The inclusion of methotrexate as a comparator vs. conventional therapy was initially confusing, but then easily understood as the literature was reviewed and practices changed over time. For Key Question 5, I would have liked the JADAS to be included as a recently published composite disease activity score, but other than that, the list was fairly inclusive and certainly appropriate.	In consultation with the TEP, the JADAS was not selected as a priority measure. This study focused on the most commonly used measures.
Peer Reviewer #7	Introduction	The introduction is well written and informative, although there is a repetition of much of the same information that is in the executive summary.	The Executive Summary is intended to be readable as a stand-alone summary of the report; therefore, some repetition is unavoidable.

Peer Reviewer #7	Introduction	Table 1 (DMARDs evaluated): I had some issues with the "mechanism of action" descriptions for some of the medications, which were ometimes inconsistent and occasionally inaccurate. For example, abatacept is described as "anti-CD28, T cell costimulator antibodies", but it is in fact not an antibody but a soluble fusion protein receptor that blocks CD28. Anakinra, canakinumab and rilonacept are all "IL-1 blockers". However, while anakinra is correctly described as a an IL-1 receptor antagonist, canakinumab is described only as an IL-1 blocker (more accurately it is an anti-IL-1beta monoclonal antibody) and rilonacept is described again as only an IL-1 blocker (a more accurate description would be a soluble fusion protein IL-1 receptor). The same is true for etanercept, adalimumab and infliximab which are all described as TNF blockers (which is true, but they have differences: etanercept is a fusion protein TNF receptor, while infliximab and adalimumab are anti-TNF monoclonal antibodies). Also, cyclosporine is described as a calcineurin inhibitor (correctly), but tacrolimus is not (when it is also a calcineurin inhibitor). So, while it doesn't really matter whether these medications are described in more general or specific terms, it should be done in a consistent manner throughout.	Changes have been made to reflect more accurate descriptions of specific mechanisms.
Levine, Loree	Introduction (Table 1)	Reviewer asked that "mechanism of action" cell for Abatacept be changed to: T-cell co-stimulation modulator; soluble fusion protein	Changes have been made.
Olfman, Joshua (Amgen)	Introduction (Pg 21)	Table 1 includes a column with a simple Yes/No entry for FDA approval for JIA. However, it is important, especially in children to define the age criteria for the labeled indication. Etanercept is approved for reducing signs and symptoms of moderately to severely active polyarticular JIA in patients 2 years of age and older. Amgen requests that the lower limit of the age range be added for abatacept (6 years), adalimumab (4 years), etanercept (2 years), methotrexate (2 years), and sulfasalazine (6 years).	This level of detail goes beyond the scope and intent of this report. However, a note has been added to each table to indicate that labeling refers to any pediatric approval.
Peer Reviewer #1	Methods	Solid	Thank you.
Peer Reviewer #3	Methods	My main concern is the limited literature search. The search was limited to PubMed and the Cochrane Database of Systematic Reviews. There is ample evidence from methods studies, however, that PubMed is not enough. It's not clear to me why other relevant electronic databases, such as EMBASE, International Pharmaceutical Abstracts or CENTRAL have not been searched. In addition, no manual literature searches appear to have been conducted.	We have now included EMBASE®.

Peer Reviewer #3	Methods	The methods of assessing the risk of bias should be more explicit. Authors state that they have modified the AHRQ guidance but these modifications have not been made transparent in the report. Particularly, because some justifications for rating individual studies as poor in the text appear to be unusual. For example, on page 39, authors state that a study on anakinra was rated as poor because of insufficient statistical power, lack of reporting of methods, and conflict of interest. All three of these issues are important but most methodologists would argue that they do not consistently affect the risk of bias in RCTs. I also think they are not listed as domains to consider in the AHRQ methods guide.	From the Methods Guide for Grading the Strength of a Body of Evidence When Comparing Medical Interventions: "Risk of bias is the degree to which the included studies for a given outcome or comparison have a high likelihood of adequate protection against bias (i.e., good internal validity), assessed through two main elements: study design … [and] aggregate quality of the studies under consideration." As suggested by the reviewer, the greatest risks of bias are related to study design and implementation. This is described in the methods. The quality assessment for key questions 1, 2, and 4 has been rewritten to reflect this approach.
Peer Reviewer #3	Methods	Authors talk about horizon scanning in the results. The rationale and the methods of this approach are not outlined in the methods chapter.	Horizon scans were not used in this report. Discussion of horizon scans has been removed.
Peer Reviewer #3	Methods (Pg 28)	Authors use the terms "attrition" and "loss to followup" when describing limitations to internal validity. What is the difference between the two?	Attrition refers to subject withdrawal. Loss to followup refers specifically to withdrawal where individuals have been lost or otherwise cannot be located.
Peer Reviewer #4	Methods	The inclusion and exclusion criteria are justifiable. Definitions and diagnostic criteria for outcome measures are appropriate within the context of the limitations of the available studies/data. Statistical methods used are appropriate.	Thank you.
Peer Reviewer #4	Methods	A couple of minor points about the search strategy. It appears the search terms for "JIA" and "JRA" were used, but not "JCA." JCA is included in the definition of the target population (KQ1, p.19). Is JCA not in the Medline nomenclature? It's stated on p. 23 that Medline was searched from 1996-2010. Why not before 1996? Or is this a mistake given that numerous studies prior to 1996 are cited and used for analysis? Similarly, only abstracts and FDA AERS data from 2008-2009 were searched. Why not for earlier years?	JCA is not a MeSH term. "1996" was an error – we searched all of MEDLINE (that is, from 1966 onward with no date restriction). This has been corrected throughout. The FDA AERS evaluation for this review was not helpful to the report and has been deleted.

Peer Reviewer #4	Methods	Looking at the list of cited studies, I have no reason to believe that the search strategy missed any studies that should have been included. However, a reader not familiar with the paucity of data about DMARD therapy in JIA may consider such a short date range for the search strategy inadequate, particularly given that the introduction implies that these medications have been in use for 25 years (p.9 line 2). Maybe a simple explanation should be included?	This was a typo and has been corrected.
Peer Reviewer #5	Methods	I was initially concerned about the search strategy (line 35, page 23) when I read the MEDLINE search was from 1996 to January 2010, and I was unsure how older DMARD studies would be captured in the search. However, older studies were referenced in the tables and in the body of evidence, so I am unclear whether they were captured in the Cochrane database reviews, or the review was actually more inclusive than stated. Selection of studies was acceptable and search strategies logical. I found that the adverse reaction reporting system was a bit confusing, reporting by systems as well as by specific symptoms. I found the system reports to be less useful. It was disappointing that the majority of the literature was analyzed qualitatively, but again, this is a consequence of the quality of studies being analyzed, not the authors or methods.	As described, that was an editing error. The adverse reaction reporting system evaluation conducted specifically for this review has been removed. We experimented with several different approaches for presenting the large amount of data pertaining to adverse events. Although differentiating between symptoms and systems is somewhat arbitrary, we believe this is preferable to listing every uniquely defined adverse event without categorization.
Peer Reviewer #6	Methods	In general, excellent. Inclusion and exclusion criteria clear except that it is not clear to me how the Brewer 1986 and Giannini 1992 papers came to be included in the review considering they are not in the interval of time specified. Excellent definitions of comparators. Outcome measures appropriate as are statistical methods. Rating the body of evidence is defined, but the reasons that evidence was rated low for some outcomes despite "good" studies is a bit unclear. Was it largely number of studies and number of patients? If so should be explicitly stated.	As described, the review began in 1966. The evidence rating strategy is now more completely described.

Peer Reviewer #7	Methods	The inclusion of JIA, JRA and JCA are certainly justifiable, as these are the main and most important terms that have been used to describe chronic arthritis in children. The search strategies were explicitly stated and logical, but I did have a question about not specifically including Juvenile Ankylosing Spondylitis and Juvenile Psoriatic Arthritis. Psoriatic arthritis is included within the umbrella term of JIA, but JIA has been in common usage only relatively recently, and these terms are not included under the JRA umbrella. Juvenile Ankylosing Spondylitis (JAS) and the JIA sub-type Enthesitis related arthritis do overlap, but there may be specific studies looking at JAS that may not have been looked at. Again, prior to the common usage of the term JIA, JAS and juvenile spondyloarthropathy or spondyloarthritis would not have been included in studies looking specifically at JRA. I do not personally know of any significant studies that looked at JAS and JPsA specifically as separate categories, so it probably will not impact the findings, but perhaps this should be looked into, and a sentence or two referring to the fact that this was done should be included. Otherwise, I thought that the data abstraction methods, quality assessment, rating the body of evidence, statistical methods, applicability assessment and data synthesis methods appeared to be appropriate and quite sound for the questions being asked.	We used inclusive and broad terms to capture JIA. These specific conditions were not separately included in the search.
Peer Reviewer #1	Results	For KQ5, the CHAQ is a disability score and is not intended to capture current disease activity – disability is often due to damage, not current inflammation. The comparisons presented are fine, but the nature of the CHAQ should be stated more clearly to avoid confusion.	The nature of the CHAQ as a disability measure has been clarified.
Peer Reviewer #2	Results (Table 4)	Table is amazing and very helpful.	Thank you.
Peer Reviewer #2	Results (Pg 39, line 45)	Why point out a potential conflict of interest due to funding for Anakinra? All of the more recent biologic studies were industry sponsored and industry analyzed. This concern should be raised for all of them or none.	As described above, industry funding is no longer specified as a risk to bias.
Peer Reviewer #2	Results	Infliximab – while the one RCT did not find a significant difference between treatment arms it was a VERY important study for several reasons that should be stated. This study established the placebo rate for children with IV treatments, it documented the best dose for infiliximab, and it reinforced that kids need their own PK and efficacy studies.	This review did not evaluate pharmacokinetics of the drugs.
Peer Reviewer #2	Results (Meta-analysis Pg 41)	It would be helpful to pull out the systemics and compare them separately (if possible)	This is a helpful suggestion. Unfortunately, the sample size precluded this analysis.

Peer Reviewer #2	Results (Figure 4)	With regard to Risk Ratio – "M-H, Fixed" is confusing and will not be understood by most readers. If it needs to stay in, please explain.	Based on the advice of another reviewer (see below), we changed to a random-effects model. This had very little impact on the analysis. We clarify in the text that this is a random-effects model. However, a description of meta-analysis goes beyond this report.
Peer Reviewer #2	Results	Methotrexate comments should be made regarding low doses used, and oral route – both of which greatly effect the efficacy of MTX.	Doses and routes for each drug are specified in the evidence table for each study. Because of the complexity of dose and routes across all drugs, we did not incorporate this information into the text of the report for methotrexate or the other drugs.
Peer Reviewer #2	Results (Table 9)	It would be important to add the consensus derived definition of remission	This is included in the section on important definitions. It is not included in the table, as the paper cited is based on consensus definition with a statement that validation studies are ongoing.
Peer Reviewer #3	Results	For the most part, more synthesis of the evidence would be helpful for readers. Right now studies are described individually, without telling the reader what the overall message of the entire body of evidence is. I also think it would be informative if the sample sizes of individual studies were made explicit in the text. Most of these trials are very small and it's important to emphasize this.	Study sample sizes are included in Tables 4, 5, and 6.
Peer Reviewer #3	Results (Figure 2)	Does not contain all the drugs listed in Table 1.	This comment refers to Figure 3. In any case, we included all studies of DMARDs that met the inclusion criteria.
Peer Reviewer #3	Results (Figure 4)	I think a random effects model would be more appropriate for a meta-analysis pooling different drugs. With a fixed effects model authors assume that all of these drugs have identical treatment effects and results of studies just vary by chance.	Change has been made. This had only a small effect on the analysis.
Peer Reviewer #3	Results (Pg 46)	Leflomide should be Leflunomide.	Change has been made.
Peer Reviewer #3	Results (Pg 52)	One reason for the low adverse events rate in the etanercept studies are the active run-in periods that excluded children who had adverse events.	We appreciate, and agree with, this observation. We have added the following sentence as a Key Point: "Adverse event rates may be underestimated by clinical trials that excluded patients who did not tolerate an intervention during a run-in phase."

Peer Reviewer #4	Results	There is an appropriate amount of detail presented. Study characteristics are adequately described. Figures and tables are readable and descriptive. It's not clear to me why the study of etanercept's effectiveness on uveitis was included (p. 40, paragraph 2). Is this relevant to the key question? And if this study is to be included, are there not other studies of DMARD use for JIA-associated uveitis that should be included? I believe studies of gold should be included. Please see my comments above.	Uveitis was included in the key questions (see Figure 1. Analytic Framework). No other studies that assessed uveitis met the inclusion criteria. We included only studies (including those of gold) if they met the a priori inclusion and exclusion criteria.
Peer Reviewer #5	Results	The results section was delightful to read. It was clear, concise, well organized by key question. Out of 2072 potential articles, only 166 articles could be included in this review. The key points were a great addition before the detailed data under each question. Explainations of how literature was graded was clear and useful. On page 52 (line 24) there appears to be a discrepancy in the report, for the 7 (42%)--This seems to be too high of a percent based on the other data in the paragraph. Additionally on page 59 (line 8) in the table it mentions 76 joints, but in the previous text on page 56, mentions active joint count is compiled from 71 joints.	We thank the reviewer for identifying these typographical errors, which we have now corrected in the body of the report. Will refer to standard joint count, with 71 joints assessed throughout.
Peer Reviewer #6	Results	Level of detail is appropriately granular. Characteristics of studies are clearly described, except I believe the statement that characterization of definition of flare was not as varied among studies as other variables (comparators, follow up periods, page 32). The statement that there were no good quality RCTs comparing biologic DMARDS to conventional therapy is based on the assumption that MTX is not conventional, yet in the introduction it says MTX is often considered part of conventional therapy, and that is how I consider it. Consider redefining conventional therapy. Figures, tables and appendices are comprehensive. No studies were overlooked, to my knowledge. In Figure 4, please define Biologic and DMARD "events" as flares. In description of leflunomide vs methotrexate study, page 44, consider adding that this was a non-inferiority study design. In adverse event analysis, page 45, why weren't open label run-in phase data in the withdrawal studies included in the analyses to make the data more robust, and not just in responders and subjects who were randomized. In measures of disease activity, page 54, the active joint count is characterized as assessing 71 joints, yet in table 8 it says 76 joints with a score from 0-73. This is confusing. Please clarify.	The discussion about the role of methotrexate has been expanded. We added a statement that the leflunomide vs. methotrexate study was a non-inferiority study design. In the adverse event analysis, we only included data from the RCT phases of multiphase studies in order to have an appropriate comparator. Clarification has been made that events refers to flares. Will define active joint count as 71 joint count since this is the most standard measure used in studies.

Source: ***www.effectivehealthcare.ahrq.gov***

Peer Reviewer #7	Results	The amount of detail was certainly appropriate, and nicely outlined, although I did have some specific comments (see below). In general, the figures, tables and appendices were more than adequate, and were done in a very detailed, consistent and sound manner, and the information was presented very clearly and nicely.	Thank you.
Peer Reviewer #7	Results (Pg 39, line 45)	Why is the fact thar there a potential conflict of interest because of the source of funding cited for this study? It is an industry sponsored clinical trial, but so are many of the other studies cited where this fact is not mentioned. There should be more consistency to applying this criticism. This brings up the question of whether each study have the funding sources listed, possibly in the tables 4 and 5. If this is a potential bias, then it is probably a good idea. The only other question is whether study sponsorship conflict of interest was used to evaluate the studies in which this was not mentioned. For consistency's sake, it does need to applied to all industry sponsored studies.	The description of industry funding has been removed.
Peer Reviewer #7	Results (Pg 39, line 57)	Says evaluated only children with polyarticular JRA, but in fact any subtype as long as it had a polyarticular course could be included, including patients with systemic JRA who no longer had systemic featuers.	This has been corrected.
Peer Reviewer #7	Results (Pg 40, line 14)	Again conflict of interest is cited because of the funding	The description of industry funding has been removed.
Peer Reviewer #7	Results (Pg 41, line 7-10)	"There was potential for significant conflict of interest because the data were analyzed by the study sponsor, which had a financial interest in tocilizumab." See above.	The description of industry funding has been removed.
Peer Reviewer #7	Results (Pg 46, lines 51 & 53)	Leflunomide is misspelled.	This has been corrected.
Peer Reviewer #7	Results (KQ 3)	The discussion focuses on rates of AEs with each DMARD, but the rate of AEs in the placebo arm patients is not always mentioned in the text summarizing the findings for each drug. Specifically discussing each study and how the rates of AEs compared, rather than doing so in only some cases would make it easier to interpret (although the numbers are in the table).	We revised the text in the body of the report to include a summary of AEs reported among patients randomized to placebo.

Peer Reviewer #7	Results (Pg 54, line 51)	In discussing the horizon scan and the reports of Hodgkins lymphoma in citation 37, there was an additional case of Hodgkin's lymphoma reported by the authors in an addendum in the same report which is not counted, and another letter by Imundo about a 4th patient with lymphoma which is also published in the same issue. These latter two cases are not discussed or counted among the JIA patients with malignancies, and Imundo's report is specifically excluded, and I am not sure why.	We have added the third case of lymphoma reported in the addendum to the cases reported by Yildirum-Toruner et al. (Yildirim-Toruner C, Kimura Y, Rabinovich E. Hodgkin's lymphoma and tumor necrosis factor inhibitors in juvenile idiopathic arthritis. J Rheumatol 2008;35(8):1680-1), which is citation 37 in the draft report. The additional case of lymphoma reported by Imundo (Imundo L, 2008, J Rheumatol 2008;35(8):1681-2) was diagnosed in a 21-year old patient; we have not, therefore, counted this case.
Peer Reviewer #7	Results (KQ 5)	One composite measure which was not discussed and specifically excluded by the TEP was the JADAS (Juvenile Arthritis Disease Activity Score). This was only recently validated (2009) and so has not been in common use, but it does include 3 of the measures (Physician global assessment, parent global assessment and active joint count), and it has the potential to be very useful because it is not a relative measure like the Pediatric ACR score (JRA core set), so I believe it should be evaluated or at least mentioned.	As described, JADAS was excluded by the TEP. We agree that this instrument is potentially very important. However, we were constrained by the recommendations of the TEP.
Levine, Loree	Results (Pg 39)	Reviewer requests the text be changed to: One good-quality randomized discontinuation study evaluated abatacept. This was a 3-phase study (Period A) was a 4-month, open-label phase that determined response to abatacept for patiente with active JIA. The second phase (Period B) was a 6-month, randomized, double-blind, parallel-dosing, placebo-controlled treatment period for all responders. Patients were given the option to receive open-label treatment with abatacept in a 5 year follow-up long term extension (Period C) if they had flare of arthritis in the double-blind period (Period B) of this study, the difference in time to disease flare between the abatacept and placebo groups was statistically significant based on the log-rank test (p=0.0002). In addition, there was statistically significant improvement compared to placebo in the active joint count (4.4 vs 6; p=0.02), CHAQ score (0.8 vs. 0.7; p=0.04), physician global assessment (14.7 vs. 12.5; p<0.01), and ACR Pediatric 90 (40% vs. 16%; p<0.01). There was no statistically significant improvement in parent/patient global assessment (17.9 vs. 23.9; p=0.70) or erythrocyte sedimentation rate (ESR; 25.1 vs. 30.7; p=0.96).	In this review, we included only those studies that compare DMARDs to other DMARDs or conventional therapy. Only the 6-month randomized discontinuation trial component of this report met the inclusion criteria. Therefore, the data from this period are specifically highlighted.

Levine, Loree	Results (Pg 52)	Reviewer requests paragraph be changed to: …preceded by a four-month, open-label lead-in phase. During the double-blind phase (Period B), serious adverse events occurred in two patients in the placebo group. One patient experiences a hematoma occurring on day 108 which was deemed unrelated to the study drug by the local investigator, while the other patient experiences both varicella and encephalitis occurring on day 27 and deemed possibly related to study drug. All resolved completely without sequelae and without study discontinuation. Acute infusional adverse events were reported in one (2%) patient in the abatacept group and two (3%) patients in the placebo group; all were either mild or moderate in intensity and none were serious. Upon continued treatment, in the open-label extension (Period C), the types of adverse events were similar in frequency and type to those seen in adult patients, except for a single patient diagnosed with multiple sclerosis while on open-label treatment (reference Orencia PI 2009)	We appreciate this suggestion, but we believe that this level of detailed analysis, if reported similarly and consistently for every DMARD and study included in this report, would detract from the aims and main findings of this report.
Olfman, Joshua (Amgen)	Results (Pg 36)	Table 4 includes the 2-year (Lovell, Giannini, Reiff, et al., 2003) follow-up of the registrational study (Lovell, Giannini, Reiff, et al., 2000) of etanercept but does not include the 4-year (Lovell, Reiff, Jones, et al., 2006) or 8-year (Lovell, Reiff, Ilowite, et al., 2008 follow-up. For transparency, Amgen requests that the publications listed be added to column 1 of the table. Amgen also requests that the follow-up durations of 4years and 8 years be added to the second-to-last column.	The followup study conducted by Lovell et al. did not include a comparator group. That study was removed from Table 4. For similar reasons, the other studies were excluded. To meet inclusion criteria, there must have been a comparator. All studies of safety were addressed in KQ3.
Olfman, Joshua (Amgen)	Results (Pg 39-40)	The section under the heading Etanercept does not state that the etanercept registrational study had an 8-year extension, during which both efficacy and safety continued to be evaluated and reported. Amgen considers this omission to be significant because of the body of data, especially safety data, collected during this time period. Amgen requests that the following sentence be added: “This trial included an 8-year extension, during which efficacy and safety continued to be evaluated and reported.”	Studies without a comparator were not included in KQ1, 2, or 4.

Olfman, Joshua (Amgen)	Results (Pg 39-40)	The section under the heading Etanercept is incomplete in addressing the disease flare experienced by some patients. Amgen requests that the following sentences be added after the third sentence of the first paragraph: "The majority of JIA patients who developed a disease flare during the double-blind component of the study and who were reintroduced to etanercept up to 4 months after discontinuation responded to etanercept in open-label studies. Most of the responding patients who continued etanercept therapy without interruption have maintained responses for up to 48 months. Data demonstrate that continuous treatment with etanercept resulted in sustained improvement in clinically important signs and symptoms for up to 8 years."	Studies without a comparator were not included in KQ1, 2, or 4.
Olfman, Joshua (Amgen)	Results (Pg 46)	The summary of the etanercept versus infliximab study includes a description of an open-label study that the authors characterize as "poor quality"; the data selected from this study tended to numerically favor infliximab, although none of the differences were statistically significant. American College of Rheumatology (ACR) Pediatric scores (other than ACR Pediatric 75 at month 12) are not included, even though the methods section of the draft report states that the authors examined ACR scores as a composite measure of disease status or response to therapy. Some of the numeric differences in ACR scores favoring etanercept were more striking (>10%) than the data that numerically favored infliximab. For balance, Amgen recommends that the following sentence be added: "After 12 months of treatment, 89% of patients treated with etanercept vs 78% of patients treated with infliximab achieved ACR Pediatric 50.	As the reviewer points out, none of the differences was statistically significant. The text describes treatment outcomes as similar. The ACR Pediatric 75 results were added as requested. No p-value was presented. We calculated a p-value for the report.
Olfman, Joshua (Amgen)	Results (Pg 46)	The section under the heading Etanercept versus Invliximab omits key information regarding the striking difference in discontinuation rates between the treatment arms. Amgen recommends that the following sentence be added: "Of the 10 patients enrolled in the etanercept arm of the trial, 1 failed to complete 12 months of therapy; that dropout was due to noncompliance. Of the 14 patients enrolled in the infliximab arm, 5 patients failed to complete 12 months – 4 because of side effects and 1 because of lack of efficacy.	This is now described.

Olfman, Joshua (Amgen)	Results (Pg 49)	Table 7 does not include safety data on adalimumab from the pivotal trial in juvenile rheumatoid arthritis (Lovell, Ruperto, Goodman, et al., 2008). For completeness, Amgen requests that the authors revise the table to include data from this trial. Some examples of imbalances in adverse event rates that favor the safety of placebo over adalimumab are: viral infection (7 reports for adalimumab vs. 3 reports for placebo; 8 reports for adalimumab + MTX vs 4 reports for placebo + MTX) and excoriation (10 reports for adalimumab vs 1 report for placebo; 6 reports for adalimumab + MTX vs. 2 reports for placebo + MTX). The adalimumab safety data should also be added under the heading Placebo-controlled RCTs of Biologic DMARDs on page 52.	We summarized the findings from Lovell et al. in the text. The adverse events data from that study are included in the Appendix. We did not include safety data on adalimumab in Table 7 because of the absence of a comparator for this purpose (adalimumab was administered in both study arms).
Olfman, Joshua (Amgen)	Results (Pg 64, paragraph 4)	The section on reliability includes a description of inter-rater reliability; the authors report that 2 studies demonstrated discordance of 60% between global assessments made by physicians vs. parents (PGA vs. PGW). Since patients and physicians perceive disease status differently, PGA and PGW were considered to be separate measures in these studies. When testing inter-rater reliability, the same measure should be used. Amgen requests that this paragraph be deleted.	The purpose of study #1 (ref 71) was to see if there was agreement between parents and physicians (the "2 raters") about whether a patient was in remission, using physician and parent global, with two raters deciding if inactive disease; the second study was to measure discordance in rating disease status. However, there was an error in the description of second study that has been corrected.
Peer Reviewer #3	Summary and Discussion (Pg 74 and following)	I like the GRADE profiles but I think they need more explanation. Not everyone is familiar with GRADE. I would explain the main concept of these tables somewhere. The GRADE BMJ series might provide good citations. I would also indicate the drugs that you are referring to. For many of the included biologics no evidence was available at all. So a moderate SoE for symptoms only applies to a few drugs not the entire class. The footnotes that GRADE tables usually have would be helpful to make the rationale for downgrading explicit.	A reference describing the GRADE profiles has been added. The goal of the summary tables is to describe the level of evidence across all classes. Specific drugs are described in the text. Methotrexate, which has a greater body of evidence, is now discussed in greater detail.
Peer Reviewer #3	Summary and Discussion	I was surprised that there was no mention of the effect of active run-in periods on applicability in trials of biologics. E.g., the etanercept study (Lovell, et al.) used an active run-in phase before randomization. They excluded all children who did not respond, had serious adverse effects, or were not adherent. So the study population was extremely selected and most likely not representative of average children treated for JIA.	We appreciate, and agree with, this observation. We have added the following sentence as a Key Point for KQ3: "Adverse event rates may be underestimated by clinical trials that excluded patients who did not tolerate an intervention during a run-in phase."
Peer Reviewer #4	Summary and Discussion	The report provides a very clear, succinct, yet thorough discussion which really emphasizes the limitations in the available data and the considerations for future studies.	Thank you.

Source: ***www.effectivehealthcare.ahrq.gov***

Peer Reviewer #4	Summary and Discussion	A statement about considering standardized outcome measures in future trials should be added to the section "Future trials in this domain should consider..." (p. 80). Unfortunately, resources for large good qulaity studies just do not exist, and therefore it's likely that studies such as the ones cited her will continue. Assuring or recommending standardization of outcome measures could improve future comparative analyses. This challenge is well stated in the previous section (p 80, lines 33-42), but not stated in the following section except regarding adverse events.	This has been added.
Peer Reviewer #4	Summary and Discussion	Additionally, "macrocyte activation syndrome" should be "macrophage activation syndrome" (p. 81, line 16 and p. 15, line 42).	We have corrected these two errors.
Peer Reviewer #5	Summary and Discussion	The discussion portion was clear and adequately stated, but as mentioned above, it is hard to deduce real conclusions based on "low" and "insufficient" evidence, which unfortunately is the strength of the majority of the evidence reported. This again illuminates the dire need for well-planned research in pediatrics, specifically as it pertains to therapeutics. This work also reveals the need for improved outcome measures from which to test therapeutic outcomes in JIA. Another large hurdle that our subspecialty is facing.	
Peer Reviewer #6	Summary and Discussion	The implications of the findings are clearly stated as are the limitations of the review. The section regarding future research is right on target, especially the large cohort studies that are suggested. The discussion may have to be modified based on my queries and suggestions in the a-d sections of this review.	No change required.
Peer Reviewer #7	Summary and Discussion	The implications of the major findings and limitations are indeed clearly stated and described through the tables presented and in the discussion about future studies. The paucity of evidence is certainly appalling, and well highlighted by this report. One limitation that is not mentioned, however, which is that the existing evidence is almost entirely centered around patients with polyarticular forms of JIA/JRA/JCA, and although in some studies other “onset” forms such as oligoarticular/pauciarticular or systemic are included, there have not been specific studies that look at specific effect of any of these medications on persistent oligoarthritis (the most common type of JIA), or on systemic arthritis specifically. The sub-types of JIA/JRA/JCA are all quite different diseases in terms of presentation, response, outcome, and probably biology, so this should be mentioned.	This has been added.
Olson, Judyann	Summary and Discussion	Would clarify the uniqueness of systemic arthritis and that mechanistically it might be different – more in the auto-inflammatory category	Clarification has been made.

Olson, Judyann	Summary and Discussion	This is an awful piece of inadequate work summarizing what are known to be bad studies. Juvenile idiopathic arthritis is not a disease. Lumping the studies together is highly unscientific and fraught with error and your conclusions are useless. Citing bad studies as insufficient may be correct, but will only encourage insurance companies to deny necessary treatment to children and result in many children suffering unnecessarily. Studies like this are not science. The literature is poor in children with arthritis and isn't worth studying. Bad conclusions based on bad literature are not worth anything.	The process used for evidence review and synthesis as described in this report is standard for the evaluation of comparative effectiveness. Identifying where evidence is insufficient is an important step in filling gaps.
Peer Reviewer #1	Future Research	Future Research section is a nice outline of our needs in JIA.	Thank you.
Peer Reviewer #7	Future Research	The suggestions regarding future trials are excellent and in general right on target. However, because JIA is a relatively rare disease with multiple sub-types, doing classical clinical trials comparing these many treatments will not be feasible. So looking at standardizing treatments, collecting uniform measures at pre-set time points and analyzing the efficacy of these treatments in an observational fashion is much more feasible, practical and likely to generate more useful information, as long as the analysis is done correctly. Likewise, the usefulness of prospective disease registries as a mechanism to study comparative effectiveness and a novel way of studying the comparative safety of these medications and the incidence of adverse events would be very important to discuss or at least mention in the conclusions.	The need for standardized outcome measures has been added. The potential role of disease registries has been added.
Peer Reviewer #7	Appendix B (B-12 & B-13)	The "numbering" of the bulleted list is incorrect	This has been corrected.
Peer Reviewer #1	General (quality of report)	Superior	Thank you.
Peer Reviewer #1	General	This is a very clear and extremely rigorous report. Unfortunately, the JIA literature is not robust enough to endure such an approach.	Thank you.
Peer Reviewer #1	General (clarity and usability)	The report is very clear. Through no fault of the authors, the usability is somewhat limited – the main message is that many rigorous studies are needed.	Thank you.
Peer Reviewer #2	General (quality of report)	Superior	Thank you.

Source: ***www.effectivehealthcare.ahrq.gov***

Peer Reviewer #2	General	This is an excellent, very thorough compilation and review of available treatment studies for JIA. It represents an enormous amount of work and the authors should be commended. All of the studies are in one place, all fairly and similarly reviewed for quality with presentation of information that can be compared (as much as possible) between studies. The tremendous value of this effort for the Pediatric Rheumatology community, patients, and families is the documentation of the profound lack of quality investigations of effective treatments in JIA. The best studies to date have been those performed by industry as part of their FDA approval process. This manuscript also points out the critical need for funding for comparative effectiveness studies in JIA and the establishment of a national registry of all JIA patients in an effort to gather meaningful long-term data on safety of the medications used to treat JIA.	Thank you.
Peer Reviewer #2	General (clarity and usability)	Very clear manuscript. Information usable to form the basis to move forward with much needed treatment studies in JIA. Points out how little information is available to help with informed decision making. Much, much more information is needed to help with policy decisions.	Thank you.
Peer Reviewer #3	General (quality of report)	Fair	See responses to individual comments from Peer Reviewer #3.
Peer Reviewer #3	General	Overall, the report is well written and for the most part logically structured. The key questions are appropriate and clinically relevant	Thank you.
Peer Reviewer #3	General	The use of the terms “efficacy” and “effectiveness” is inconsistent throughout the report. Authors distinguish the terms in the key questions but often use the term “effectiveness” in the text or in table A when “efficacy” would be the appropriate term.	This has been clarified.
Peer Reviewer #3	General (clarity and usability)	Overall, I think the report is well structured. The only part where I have been struggling with the structure has been KQ3. The way it is structured now (by drug) makes it really hard to understand what the main safety concerns of these drugs are and whether there are any patterns that might indicate differences in harms. I think the section would be more readable and informative if it were structured by drug class (e.g., biologic, DMARDS, synthetic DMARDS, etc.) and then by harms. E.g., for biologics: infections, cancer, hematological disorders, infusion/injection site reactions, etc. Within the specific harms the evidence on the various drugs could be summarized. At the moment one has to go back and forth among drug paragraphs to get some idea of the comparative harms.	We gave careful consideration to this suggestion. We decided, in the end, to keep the same structure and organization across all key questions.

Peer Reviewer #3	General (clarity and usability)	In addition, I think the report could be improved if key points contained more information on differences among drugs, the magnitudes of effects, and the risks of harms. E.g., statements like "there is some evidence that methotrexate is superior to conventional therapies…" are vague and do not convey anything about treatment effects.	Overall magnitude of effect is difficult to describe. However, the description of methotrexate has been expanded.
Peer Reviewer #4	General (quality of report)	Superior	Thank you.
Peer Reviewer #4	General	This is a very exhaustive and well done comparative effectiveness review on a challenging topic. Whether or not this report is "clinically meaningful" depends on what one means by "clinically meaningful." The results of this analysis are unlikely to change how I care for patients with JIA, but that is not because of any particular limitation of this report. It is because of the limitations of the available data, limitations which this report very thoroughly describes. If anything, this report points out the true paucity of quality data supporting the effectiveness and safety of DMARDS in treating JIA, which may challenge the current thinking regarding their safety and effectiveness, but is unlikely to significantly change their use, because they are "all we know" and "all we have." The real strength of this report is its ability, through thorough analysis, to make explicit recommendations for future research. And in that regard is very clinically meaningful.	Thank you.
Peer Reviewer #4	General (clarity and usability)	This report is well structured and clearly presented. The conclusions are certainly meaningful when it comes to guiding future research, policy, and resource allocation. Practice decisions may also be influenced, but less so as I suspect that most of the conclusions will not be surprising to practicing pediatric rheumatologists.	Thank you.
Peer Reviewer #5	General (quality of report)	Superior	Thank you.
Peer Reviewer #5	General	The target population and audience are explicitly defined and the key questions are appropriate and clearly stated. As far as clinically meaningful, this is more difficult. It is necessary to see the paucity of quality data from which we work, but this review is unable to come to firm conclusions based on the fact that the data from which the group worked were sparse and studies were often poorly conducted. This, however, is an initial step in illuminating the need for quality research in our field. It was well orchestrated and very clearly planned.	Thank you.

Peer Reviewer #5	General (clarity and usability)	This report is beautifully organized and clearly presented. Although the conclusions are less likely to guide therapy based on the lack of strong evidence either way, it is a foundation from which to grow and improve. It makes very clear the work that needs to be done in planning well designed therapeutic studies, both investigator-initiated and industry-initiated which will include pediatrics as a group in their analyses. This report also highlights the limitations we face in studying a rare population with a very heterogeneous phenotype, with some suggestions on how to tackle this. I see this report as a necessary foundation from which to improve the quality of the research we perform in pediatric rheumatology, with a goal for unified validated outcome measures and studies that will improve the quality of care that we ultimately provide to our patients.	Thank you.
Peer Reviewer #6	General (quality of report)	Superior	Thank you.
Peer Reviewer #6	General	In general, this report is thorough, comprehensive, clear and concise given the scope of work required for this ambitious project. The key questions are appropriate and explicit and I found the figures very helpful in understanding the analyses, particularly figure 3.	Thank you.
Peer Reviewer #6	General	I did find the definition of "conventional therapy" confusing which was probably unavoidable given that it has changed over the course of time included in the review.	The inclusion of methotrexate as both DMARD and conventional therapy is complicated. However, we believe that the target audience of this report will understand this perhaps unavoidable nuance.
Peer Reviewer #6	General (clarity and usability)	The report is well structured, organized and main points are clear. The conclusions inform future studies more than practice decisions currently, given the criticisms of many of the studies, but is appropriate.	Thank you.
Peer Reviewer #7	General (quality of report)	Superior	Thank you.

Peer Reviewer #7	General	This report is a systematic review of the evidence on the effectiveness of disease modifying anti-rheumatic drugs in children with Juvenile Idiopathic Arthritis. The topic is without question clinically meaningful, especially in light of the many new treatments that have become available over the last 10 years, and the authors have very thoroughly and exhaustively reviewed and evaluated the available information. The key questions are well thought out as are the relevant questions that need to be addressed regarding the subject, which are stated in a clear and explicit manner. The authors have done an excellent job of reviewing the relevant material, digesting the evidence regarding the effectiveness of DMARDs in JIA, and presenting the information in a clear and concise fashion. The report is striking in that it highlights the paucity of evidence regarding this subject, the poor quality of much of the evidence that does exist, and points to a real gap in knowledge in our field which is very important indeed. This report is very timely and will likely result in future work that will hopefully close this large evidence gap.	Thank you.
Peer Reviewer #7	General (clarity and usability)	The report is well structured and organized, and although there is some repetition of the information and salient points, the main points are clearly presented, and the conclusions can certainly be used to inform policy, and eventually practice decisions, once the right kinds of studies are done and enough data is generated. The fact that there is little usable information about the effectiveness and safety of many of the treatments we use for patients with JIA comes across strongly; as does the striking need for studies and new approaches designed to answer these questions about the treatments that are used. The authors should be commended for an excellent report.	Thank you.